Eddieb Sadat
Professor Zadbood
EM622-WS Decision Making via Data Analysis
11/23/22

Homework 4

Introduction:

Finance is one industry that heavily utilizes data analysis, and heavily employs the use of visualization for representation of often complex or long datasets. Examples of these can be found looking no further than the stock market, which use time series graphics to visualize the changes of stock prices (among other financial metrics) over time. These visuals have proven to be important for this industry, as it can allow analysts to more effectively identify patterns in the market, which can heavily influence financial decisions. Thus, being able to utilize time series is a key skill. To practice this skill, the closing stock price of three companies over a one year period will be analyzed, graphed, and formatted in a manner that is easy to understand.

Business Understanding:

This project is being made to complete the Homework 4 assignment for EM-622, which aims to provide opportunity for hands-on experience in time series by generating a plot of stock prices. The requirements to successfully complete the assignment are to download financial data from three companies (excluding Amazon and Google) over a one-year time span, plot them, and automatically generate labels for the minimum and maximum closing stock prices for each company. Thus, the first step for this project is to determine which three stocks would be interesting to analyze. Personally, I find the oil and energy sector an interesting industry, since it can be somewhat of an indicator for the state of the world economy. I decided to select among the largest of oil companies: Exxon, Chevron, and Shell.

As mentioned earlier, the primary objective of this assignment is to give first-hand experience in financial data by utilizing time series. However, from an industry standpoint, being able to generate effective visuals of financial data can vastly improve financial understanding and business operations for a company. Thus, this project has very real and important implications beyond the scope of the EM-622 course.

Data Understanding:

Luckily, financial data for large companies are readily available, especially in R, a language that is heavily utilized in the financial industry. As such, all financial information is

easily accessible directly in R. Using the tseries library, the closing stock price for Exxon, Chevron and Shell from January 1, 2021 to January 1, 2022 were generated.

exxon	252 obs. of 2 variables
chevron	252 obs. of 2 variables
shellplc	252 obs. of 2 variables

From the figure above, all three companies had equal values generated, which is a very good sign. It is also important to note that stock prices only change when the market is open, which excludes weekends and holidays. Thus, it makes sense that there are only 252 rows and not 365.

	Close
2021-01-04	41.50
2021-01-05	43.50
2021-01-06	44.61
2021-01-07	44.96

Data from each company was saved into a dataframe, which contains two columns: the 'name' column which contains the date, and the 'Close' column which contains the closing stock price. The data is very straightforward, and seeing as there are no missing values from any of the companies (proven by the fact that they all have the same date values), then the next step is to restructure the data to be used for visualization.

Data Preparation:

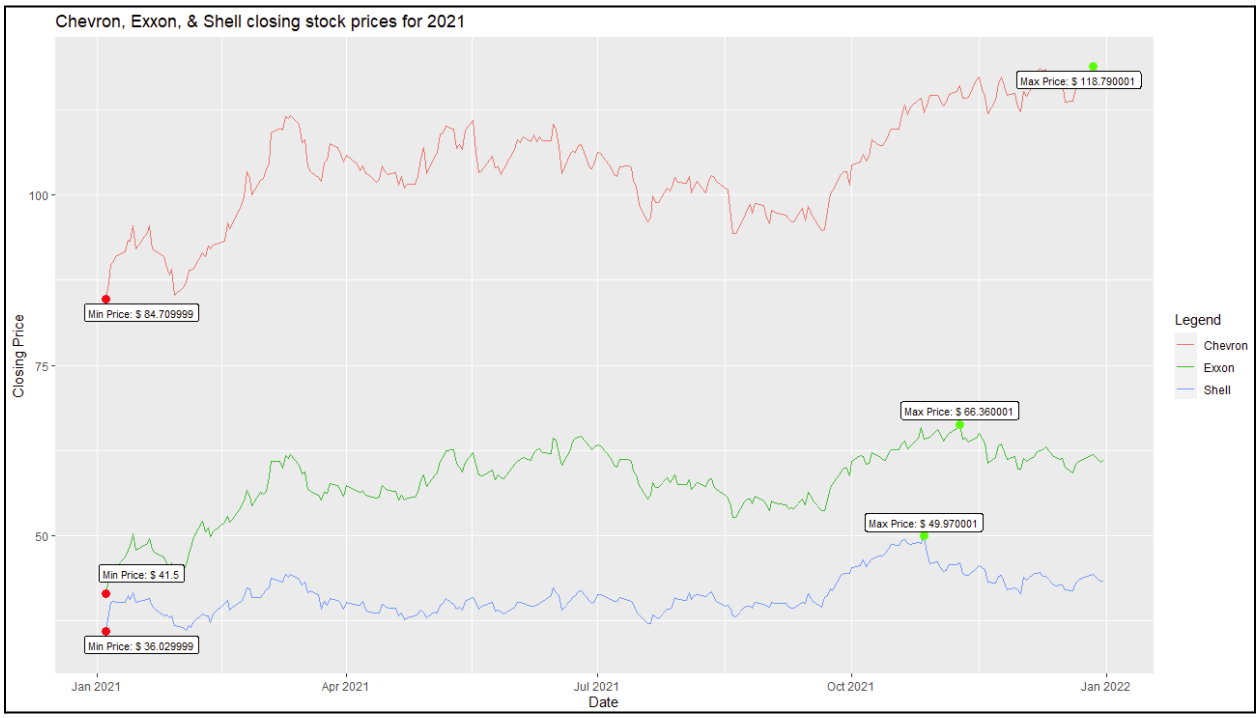
The date information for each closing price needs to be readily accessible for graphing. As such, the values from the 'name' column are pulled and saved into a new column called 'date'.

	Close	date
2021-01-04	41.50	2021-01-04
2021-01-05	43.50	2021-01-05
2021-01-06	44.61	2021-01-06
2021-01-07	44.96	2021-01-07

With this complete, the data is finally ready to be visualized.

Modeling:

Because the date is easily accessible and is already in chronological order and in equal discretization, a line plot will perfectly represent the data over time. At first I attempted to follow the suggestions given in canvas by combining all datasets into one to allow each company lineplot to be colored differently. However, I found it difficult to add the max and min labels using this method. Instead, I opted to generate three separate line plots using the three individual company datasets. With this method, I was able to manually generate a legend of colors for each company[1], and create points for the maximum and minimum closing prices for each company[2]. Finally, I was able to add labels which indicate the max and min prices[3], altered the title and axis labels, and generated the final graph.



Evaluation:

The final graph produced is very useful, as it not only shows information about each company, but it also compares them to each other. For example, it is clear that the highest stock price is Chevron, followed by Exxon and Shell. Of course this doesn't indicate which company is worth more, since that requires more information, such as the number of total stocks.

There is a very clear correlation between each company, or perhaps more accurately, they are each heavily correlated with the same thing (we can't identify this through this information alone). Each company has very similar stock growth/decay over the same time periods and their mins/maxes occur within very similar timespans, which makes complete sense considering they

are three very dominant oil/energy companies (changes in the oil market will affect all of them in fairly similar ways).

Deployment:

I am confident that the methods employed in generating this visual are perfectly capable of satisfying industry standards (maybe not the code itself… admittedly it is quite messy). The visual is not only useful in seeing individual company stock prices over time, it also allows for the easy comparison of each company. Even through this simple example important patterns were observed, and if implemented on a larger scale, can have the potential to give a great wealth of new information to stakeholders. However, in order to truly satisfy stakeholders, there must be additional financial information included that can shed more light into how financially healthy each company is.

Sources:
[1] https://community.rstudio.com/t/adding-manual-legend-to-ggplot2/41651/5
[2] https://www.r-bloggers.com/2010/02/r-tip-finding-the-location-of-minimum-and-maximums/
[3] https://www.tutorialkart.com/r-tutorial/concatenate-two-or-more-strings-in-r/